A Statistician's Journey to LLNL

Livermore WiDS

March 8, 2021

Ana Kupresanin
Deputy Division Leader
Center for Applied Scientific Computing

LLNL-PRES-820259
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

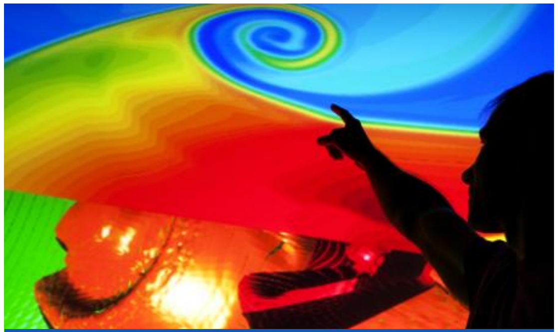
Lawrence Livermore National Laboratory

There are many great parts of my job

CASC
Center for Applied
Scientific Computing

DSI
DATA SCIENCE
INSTITUTE

PSAAP III



Weapon Physics and Design

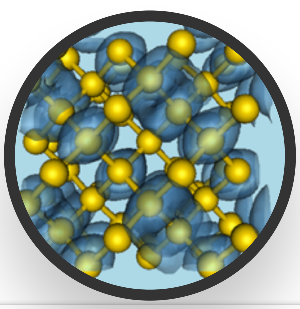
Center for Applied Scientific Computing

CASC
Center for Applied Scientific Computing

- Formed in 1996 (during early days of ASCI)
- Started with 15, more than 150 staff members today

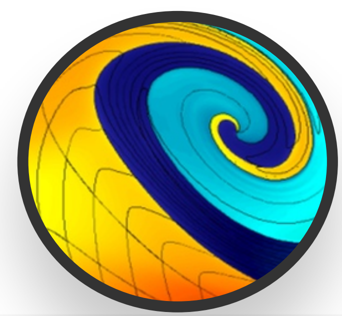
We deliver innovative research and software technologies that advance computational sciences in the national interest

Our vision is to transform applied science through mathematics, computer science, and data science research

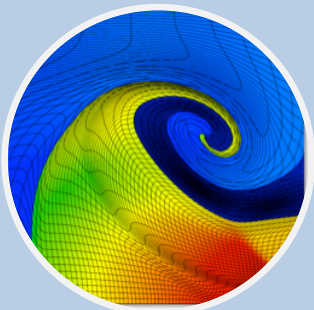


We conduct research across the breadth of computational sciences

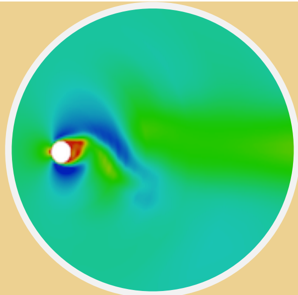
- Extreme scale algorithms and tools
- Computational Mathematics
- Data analytics and visualization, Machine Learning
- Advancing physics understanding



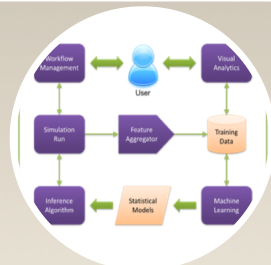
Data science and exascale computing provide an opportunity to rethink our algorithms and mathematics



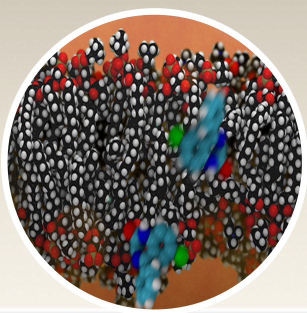
ALE Hydrodynamics
- **BLAST** code, **MFEM** lib
- High-order finite elements and meshes
- Symmetry and energy



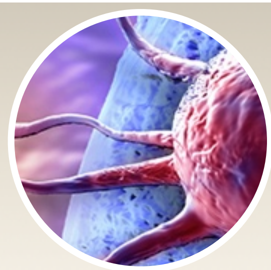
Parallel-in-Time
- **XBraid** library
- Addresses serial time integration bottleneck
- Concurrently compute multiple time steps
- O(N) multilevel algorithm



Scientific Workflow
- Enabling ecosystem of tools to construct workflows, manage simulation data, and apply next-gen analytics
- **Siboka:** Data & workflow management
- **C2C:** Problem setup
- **ROVER:** In-situ diagnostics



Learning Cancer Mechanisms
- RAS pathway problem
- Unsupervised learning coupled to multiscale molecular dynamics simulations
- Understand RAS protein interactions with membranes



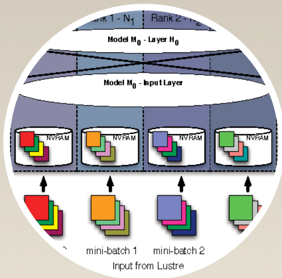
Precision Medicine for Cancer

- Revisit methods traditionally considered too computationally expensive
- Find new sources of concurrency
- Develop new models, UQ, optimization, inverse problems
- Improve workflows that couple experiment, simulation, and theory — accelerating scientific discovery
- Requires close collaboration between mathematicians, computer scientists, data scientists and applications

Use of ML and AI in mission-critical settings places new concerns and demands on these algorithms

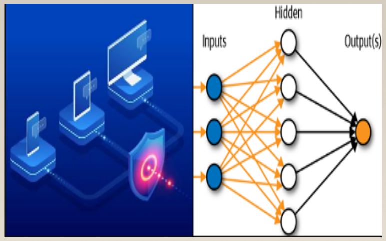
How to?

- **Make sure that the collected data is relevant and sufficient for the problem at hand**
- Select appropriate ML model form and metric
- Include existing knowledge into ML algorithms
- Assess their robustness and predictive power



Extreme-Scale Learning

- **LBANN:** Livermore Big Artificial Neural Network
- Toolkit for scalable parallel training of DNNs
- Livermore Brain project – unsupervised training of 15-billion parameter CNN on 100 million images



Assured ML

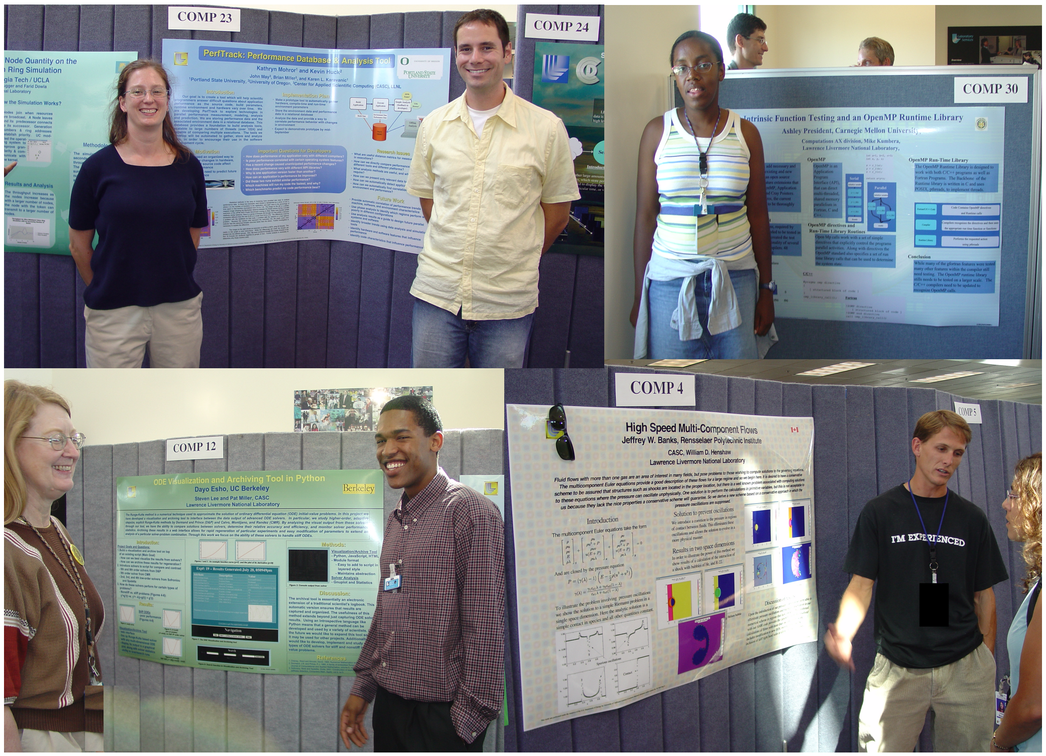
- Enable predictable behavior & assurance for ML systems
- Formally prove that ML models are consistent with domain knowledge
- Train ML models that are accurate and also specification consistent

Interpretable ML

- Model introspection through example trees
- Two-stage training motivated by neuroscience
- Feed-forward learning of representations & feed-backward learning of tasks at the tail end

CASC researchers work with the scientific community to build a firm theoretical foundation for ML and AI

CASC summer internship program is integral to our identity



- Internships introduce students to work at CASC and LLNL
- Students bring fresh ideas and perspectives
- CASC contributes to training the next generation of computational scientists
- Students often return and become CASC postdocs (and later, staff)

Predictive Science Academic Alliance Program (PSAAP)

"For more than 20 years, the PSAAP program has been an invaluable resource for bringing some of the country's most promising students together with our scientists and engineers."
— Dr. Kim Budil, LLNL Director

- Engages academic community in advancing science-based modeling and simulation technologies
- Managed by the NNSA Office of Advanced Simulation and Computing in collaboration with LLNL, LANL and SNL
- LLNL provides supercomputing resources to the nine PSAAP III centers
- https://psaap.llnl.gov

"The promise of greatly enhanced productivity through the use of artificial intelligence, enabled with heterogeneous architectures, is something we see both our staff and PSAAP students to be highly energized over." — Dr. Chris Clouse, ASC Director

Nine PSAAP centers support modeling and simulation capabilities relevant to stockpile stewardship

University of Texas at Austin: Exascale Predictive Simulation of Inductively Coupled Plasma Torches

University of Colorado at Boulder: Micromorphic Multiphysics Porous and Particulate Materials Simulations with Exascale Computing Workflows

Stanford University: Integrated Simulations using Exascale Multiphysics Ensembles

University of Illinois: Exascale-Enabled Scramjet Design

Massachusetts Institute of Technology: Exascale Simulation of Material Interfaces in Extreme Environments

University at Buffalo: Exascale Simulation of Hybrid Rocket Motors

University of New Mexico: Understandable,Performant Exascale Communication Systems

University of Maryland: Solution Verification, Grid Adaption and Uncertainty Quantification for Chaotic Turbulent Flow Problems

Oregon State University: Exascale Monte Carlo Neutron Transport

PSAAP prepares the next generation of scientists and engineers for careers in national security

- Completing an internship at an NNSA lab is part of students' curriculum
- Internship exposes the student to a collaborative interdisciplinary team working environment
- PSAAP provides an avenue for LLNL scientists to interact with academia
  - More than 20 LLNL staff members are associated with PSAAP centers through serving on advisory teams or reviewing centers' progress

Data Science Institute

- Established in 2018
- https://data-science.llnl.gov/
- Integrating organization focused on growing, strengthening, and sustaining LLNL's data science workforce
- DSI addresses rapid growth of data science and its impact on the Lab's mission

"LLNL is a data-rich environment. As the demand for sophisticated methods of analyzing and interpreting data grows, so too does the need to push the boundaries of data science"
— Dr. Bruce Hendrickson
LLNL associate director for Computing

DSI is a team effort and includes representatives from all disciplines

Data Science Council

Peer-Timo Bremer
Comp

Ana Kupresanin
Comp

Dan Faissol
Eng

Barry Chen
Eng

Michael Schneider
PLS

Organizational Chart

Mike Goldman, Director
Comp

Jen Bellig, Admin
Comp

Data Science Ambassadors

David Buttler, **Comp**
Rushil Anirudh, **Comp**
Jason Bernstein, **Eng**

Communications & Web

Holly Auten, **TID**
Mary Gines, **TID**
Kyle McAneney, **TID**

Data Science Summer Institute

Goran Konjevod, Director
Eng

Nisha Mulakken, Co-director
Comp

Jennie Carranza, Admin
Comp

DSI is creating a thriving collaborative research community internally and externally

- DSI Seminar Series
- Open Data Initiative https://data-science.llnl.gov/open-data-initiative
- DSI Consulting Services
- External partnership with University of California system
- Annual DSI workshop
- Reading groups
- Faculty summer visits
- Data Science Challenge

DATA SCIENCE SANTA CRUZ

Data Science Initiative UCDAVIS UNIVERSITY OF CALIFORNIA

UC San Diego HALICIOĞLU DATA SCIENCE INSTITUTE

UNIVERSITY OF CALIFORNIA UCMERCED

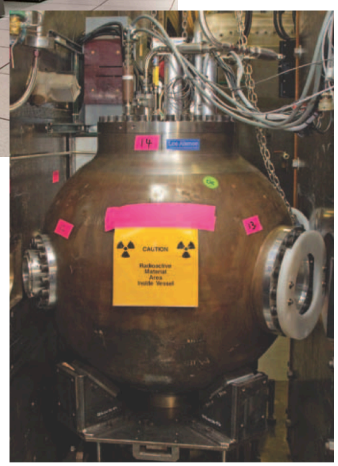
UNIVERSITY OF CALIFORNIA UCRIVERSIDE

BIDS BERKELEY INSTITUTE FOR DATA SCIENCE

UC SANTA BARBARA | Data Science Initiative

My journey to LLNL

Our L1 milestone team made a blind prediction with quantified uncertainties of a LANL experiment



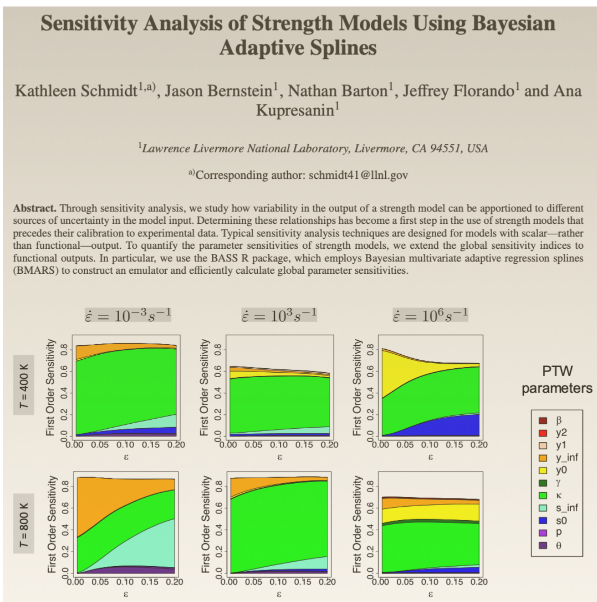
The Castor subcritical test—the test device was placed within the containment vessel
Source: https://www.lanl.gov/discover/publications/1663/2014-august/critical-subcritical.php

Our team received an NNSA Defense Programs Award of Excellence for work performed in 2012

"The team developed and exercised a functional UQ methodology that incorporated both focused and integral physics experiments. This adaptive methodology allowed the team to complete UQ studies in a fraction of the time of past methods. This achievement enabled, for the first time, a systematic analysis of all relevant focused physics experiments related to early phase hydrodynamics and how they inform our science-based models. The effort clearly described the weaknesses and strengths of the various models and where improvements need to be made."
Source: http:/str.llnl.gov/july-2014/awardsjuly14

Uncertainty quantification and novel statistical methods is an area of scientific inquiry within WPD program

We developed a method for sensitivity analysis of complex physics models for quantities of interest that change over time (or space)



Sensitivity Analysis of Strength Models Using Bayesian Adaptive Splines

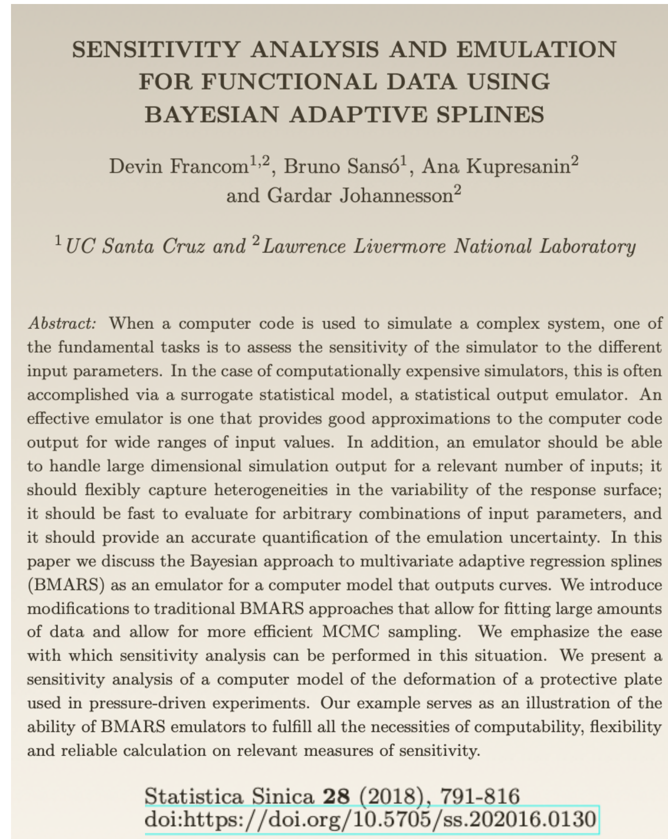
Kathleen Schmidt[1,a)], Jason Bernstein[1], Nathan Barton[1], Jeffrey Florando[1] and Ana Kupresanin[1]

[1]*Lawrence Livermore National Laboratory, Livermore, CA 94551, USA*

[a)]Corresponding author: schmidt41@llnl.gov

Abstract. Through sensitivity analysis, we study how variability in the output of a strength model can be apportioned to different sources of uncertainty in the model input. Determining these relationships has become a first step in the use of strength models that precedes their calibration to experimental data. Typical sensitivity analysis techniques are designed for models with scalar—rather than functional—output. To quantify the parameter sensitivities of strength models, we extend the global sensitivity indices to functional outputs. In particular, we use the BASS R package, which employs Bayesian multivariate adaptive regression splines (BMARS) to construct an emulator and efficiently calculate global parameter sensitivities.

AIP Conference Proceedings, 1979(1), pp. 140004, 2018



SENSITIVITY ANALYSIS AND EMULATION FOR FUNCTIONAL DATA USING BAYESIAN ADAPTIVE SPLINES

Devin Francom[1,2], Bruno Sansó[1], Ana Kupresanin[2] and Gardar Johannesson[2]

[1]*UC Santa Cruz and* [2]*Lawrence Livermore National Laboratory*

Abstract: When a computer code is used to simulate a complex system, one of the fundamental tasks is to assess the sensitivity of the simulator to the different input parameters. In the case of computationally expensive simulators, this is often accomplished via a surrogate statistical model, a statistical output emulator. An effective emulator is one that provides good approximations to the computer code output for wide ranges of input values. In addition, an emulator should be able to handle large dimensional simulation output for a relevant number of inputs; it should flexibly capture heterogeneities in the variability of the response surface; it should be fast to evaluate for arbitrary combinations of input parameters, and it should provide an accurate quantification of the emulation uncertainty. In this paper we discuss the Bayesian approach to multivariate adaptive regression splines (BMARS) as an emulator for a computer model that outputs curves. We introduce modifications to traditional BMARS approaches that allow for fitting large amounts of data and allow for more efficient MCMC sampling. We emphasize the ease with which sensitivity analysis can be performed in this situation. We present a sensitivity analysis of a computer model of the deformation of a protective plate used in pressure-driven experiments. Our example serves as an illustration of the ability of BMARS emulators to fulfill all the necessities of computability, flexibility and reliable calculation on relevant measures of sensitivity.

Statistica Sinica **28** (2018), 791-816
doi:https://doi.org/10.5705/ss.202016.0130

CASC

Center for Applied
Scientific Computing

Lawrence Livermore
National Laboratory